Outline of the course

• Introduction about Compara
  - Resources
  - API

• Base objects
  - Genes, peptides, RNAs
  - Multiple / pairwise alignments

• Data objects
  - Families
  - Gene trees
  - Homologies
**AlignedMemberSet object**

- Base object that represents a set of members aligned together, e.g. a multiple alignment of peptides / ncRNAs
- “Applied” in gene trees, families, and homologies
- No specific adaptor

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of members</td>
<td>$aln-&gt;get_all_Members()</td>
</tr>
<tr>
<td>Alignment (BioPerl object)</td>
<td>$aln-&gt;get_SimpleAlign()</td>
</tr>
<tr>
<td>Description (if available)</td>
<td>$aln-&gt;description()</td>
</tr>
<tr>
<td>Stable ID (if available)</td>
<td>$aln-&gt;stable_id()</td>
</tr>
</tbody>
</table>
HOWTO: print a BioPerl alignment

- Compara objects return alignments as BioPerl instances
  
  $aln->get_SimpleAlign()

- BioPerl provides an AlignIO object to format the actual output in various formats (fasta, clustalw, phylip ...)

```perl
use Bio::AlignIO;

# Get the alignIO object from BioPerl
my $alignIO = Bio::AlignIO->newFh(-format => "fasta");

# Print the alignment
print $alignIO $aln;
```
Outline of the course

• Introduction about Compara
  - Resources
  - API

• Base objects
  - Genes, peptides, RNAs
  - Multiple / pairwise alignments

• Data objects
  - Families
  - Gene trees
  - Homologies
Families are clusters of similar peptides

1. Ensembl (all transcripts)
2. Uniprot/SWISSPROT
3. Uniprot/SPTREMBL

All vs. all BlastP

MCL (Markov CLustering algorithm)

Mafft multiple alignment

Mafft

Mafft

Mafft
Example on the web:
ENSFM005000000271501 in Zebrafish

ZEBRAFISH genes in this family

Ensembl genes containing proteins in family ENSFM005000000271501

<table>
<thead>
<tr>
<th>Gene ID and Location</th>
<th>Gene Name</th>
<th>Description (if known)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSDARG00000041086</td>
<td>CABZ01071177.1</td>
<td>Uncharacterized protein [Source: UniProtKB/Trembl; Acc: F1QUF1]</td>
</tr>
<tr>
<td>Chromosome 7: 75.50m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSDARG00000024771</td>
<td>slc24a5</td>
<td>Solute carrier family 24, member 5 [Source: ZFIN; Acc: ZDB-GENE-031210-1]</td>
</tr>
<tr>
<td>Chromosome 11: 4.64m</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Family object / FamilyAdaptor

- Represents a group of similar peptide members

  $family_adaptor->fetch_all_by_Member(...)
  $family_adaptor->fetch_by_stable_id(...)

- Alternative transcripts can belong to different families! Families contain both SeqMembers and GeneMembers

- (almost) the same methods as in AlignedMemberSet

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment</td>
<td>$family-&gt;get_SimpleAlign()</td>
</tr>
</tbody>
</table>
| Biological function | $family->description()  
                  | $family->description_score()             |
| Gene content   | $family->get_all_Members()                   |
Exercises - Families

- Get the multiple alignment corresponding to the family with the stable id ENSFM00250000006121

- Get the families predicted for the human gene ENSG00000139618. What do you notice?